Finiteness and Fluctuations in Growing Networks

P. L. Krapivsky and S. Redner
Center for BioDynamics, Center for Polymer Studies, and Department of Physics, Boston University, Boston, MA, 02215

We study the role of finiteness and fluctuations about average quantities for basic structural properties of growing networks. We first determine the exact degree distribution of finite networks by generating function approaches. The resulting distributions exhibit an unusual finite-size scaling behavior and they are also sensitive to the initial conditions. We argue that fluctuations in the number of nodes of degree k become Gaussian for fixed degree as the size of the network diverges. We also characterize the fluctuations between different realizations of the network in terms of higher moments of the degree distribution.

PACS numbers: 02.50.Cw, 05.40.-a, 05.50.+q, 87.18.Sn

I. INTRODUCTION

Networks such as the Internet and the World-Wide Web do not grow in an orderly manner. For example, the Web is created by the uncoordinated effort of millions of users and thus lacks an engineered architecture. Although such networks are complex in structure [1,2], their large size is a simplifying feature, and for infinitely large networks the rate equation approach [3] provides analytical predictions for basic network characteristics. Nevertheless, social and technological networks are not large in a thermodynamic sense (*e.g.*, the number of molecules in a glass of water vastly exceeds the number of routers in the Internet). Thus fluctuations in network properties can be expected to play a more prominent role than in thermodynamic systems [4]. Additionally, extreme properties, such as the degree the node with the most links in a network [5,6], the website with the most hyperlinks, or the wealth of the richest person in a society, are important characteristics of finite systems. The size dependence of these properties or their distribution is difficult to treat within a rate equation approach.

In this paper, we examine the role of finiteness and the nature of fluctuations about mean values for large, but finite growing networks. We shall focus primarily on the degree distribution $N_k(N)$, the number of nodes that are linked to k other nodes in a network of N links, as well as related local structural characteristics. We shall argue that self averaging holds for the degree distribution, so that the random variables $N_k(N)$ become sharply peaked about their average values in the $N \to \infty$ limit. We shall also argue that the probability distribution for the number of nodes of fixed degree, $P(N_k, N)$, is generally a Gaussian, with fluctuations that vanish as $N \to \infty$. On the other hand, higher moments of the degree distribution do not self average. This loss of self-averaging ultimately stems from the power-law tail in the degree distribution itself.

In the next section, we define the growing network model and briefly review the behavior of the average degree distribution in the thermodynamic $N \to \infty$ limit. We also discuss how the average degree distribution can naturally be expected to attain a finite-size scaling form for large but finite N. We then describe our general strategy for studying fluctuations in these growing networks. In Sec. III, we outline our simulational approach and present data for the average degree distribution. In the following two sections, we examine the role of finiteness on the degree distribution, both within a continuous formulation based on the rate equations (Sec. IV), and an exact discrete approach (Sec. V). The former approach is the one that is conventionally applied to study the kinetics of evolving sustems, such as growing networks. While this approach has the advantage of simplicity and it provides an accurate description for the degree distribution in an appropriate degree range, it is quantitatively inaccurate in the large degree limit. This is the domain where discreteness effects play an important role and the exact discrete recursion relations for the evolution of the degree distribution are needed to fully account it properties. In Sec. VI, we discuss the implications of our results for higher moments of the degree distribution and their associated fluctuations. Sec. VII provides conclusions and some perspectives. Calculational details are given in the appendices.

II. STATEMENT OF THE PROBLEM

The growing networks considered in this work are built by adding nodes to the network one at a time according to the rule that each new node attaches to a single previous node with a rate proportional to A_k, where k is the degree of the target node. We investigate the class of models in which $A_k = k + \lambda$, where $\lambda > -1$, but is otherwise arbitrary. The general situation of $-1 < \lambda < \infty$, corresponds to linear preferential attachment, but with an additive shift λ in the rate. This model was originally introduced by Simon to account for the word frequency distribution [7]. The case $\lambda = 0$ corresponds to the Barabási-Albert model [8], while the limit $\lambda \to \infty$ corresponds to random attachment in which each node has an equal probability of attracting a connection from the new node. Thus by varying λ, we can tune the relative importance of popularity in the attachment rate.

Previous work on the structure of such networks was

primarily concerned with the configuration-averaged degree distribution $\langle N_k(N)\rangle$, where the angle brackets denote an average over all realizations of the growth process for an ensemble of networks with the same initial condition. Additionally, most studies focused on the tail region where k is much smaller than any other scale in the system. For attachment rate $A_k = k+\lambda$, this average degree distribution has a power-law tail [7,9],

$$\langle N_k(N)\rangle = N\, n_k, \quad \text{with} \quad n_k \propto k^{-(3+\lambda)} \tag{1}$$

as $N \to \infty$. In the specific case of $A_k = k$, the average degree distribution explicitly is [7–11]

$$\langle N_k(N)\rangle = N\, n_k, \quad \text{with} \quad n_k = \frac{4}{k(k+1)(k+2)}. \tag{2}$$

For finite N, however, the degree distribution must eventually deviate from these predictions because the maximal degree cannot exceed N. To establish the range of applicability of Eqs. (1), we estimate the magnitude of the largest degree in the network, $k_{\max}$ by the extreme statistics criterion $\sum_{k\geq k_{\max}} \langle N_k(N)\rangle \approx 1$. This yields $k_{\max} \propto N^{1/(2+\lambda)}$. We therefore anticipate that the average degree distribution will deviate from Eq. (1) when k becomes of the order of $k_{\max}$. The existence of a maximal degree also suggests that the average degree distribution should attain a finite-size scaling form

$$\langle N_k(N)\rangle \simeq N n_k F(\xi), \qquad \xi = k/k_{\max}. \tag{3}$$

Some aspects of these finite-size corrections were recently studied in Refs. [12–15]. One basic result of our work is that we can compute the scaling function explicitly. We find that this function is peaked for k of the order of $k_{\max}$ and that it depends substantially on the initial condition. In contrast, the small-degree tail of the distribution – the reason why such networks were dubbed scale-free – is independent of N and the initial condition.

To study finite networks where fluctuations can be significant, we need a stochastic approach rather than a deterministic rate equation formulation. For finite N, the state of a network is generally characterized by the set $\mathbf{N} = \{N_1, N_2, \ldots\}$ that occurs with probability $P(\mathbf{N})$. The network state $\mathbf{N}$ evolves by the following processes:

$$\begin{aligned}(N_1, N_2) &\to (N_1, N_2+1),\\ (N_1, N_k, N_{k+1}) &\to (N_1+1, N_k-1, N_{k+1}+1).\end{aligned}$$

The first process corresponds to the new node attaching to an existing node of degree 1; in this case, the number of nodes of degree 1 does not change while the number of nodes of degree 2 increases by 1. The second line accounts for the new node attaching to a node of degree $k>1$.

From these processes, it is straightforward, in principle, to write the master equation for the joint probability distribution $P(\mathbf{N})$. It turns out that correlation functions of a given order are coupled only to correlation functions of the same and lower orders. Thus we do not need to invoke factorization (as in kinetic theory) and we could, in principle, solve for correlation functions recursively. However, this would provide much more information than is of practical interest. Typically we are interested in the degree distribution, or perhaps two-body correlations functions of the form $\langle N_i\, N_j\rangle$. Even though straightforward in principle, it is difficult to compute even the two-point correlation functions $\langle N_i\, N_j\rangle$ for general i and j. In this work, we shall restrict ourselves to the specific (and simpler) examples of $\langle N_1^2\rangle$, $\langle N_1\, N_2\rangle$, and $\langle N_2^2\rangle$. We will use these results to help characterize fluctuations in finite networks.

III. SIMULATION METHOD AND DATA

To simulate a network with attachment rate $A_k = k+\lambda$ efficiently, we exploit an equivalence to the growing network with re-direction (GNR) [9]. In the GNR, a newly-introduced node $\mathbf{n}$ selects an earlier "target" node $\mathbf{x}$ *uniformly*. With probability $1-r$, a link from $\mathbf{n}$ to $\mathbf{x}$ is created. However, with probability r, the link is *re-directed* to the ancestor node $\mathbf{y}$ of node $\mathbf{x}$ (Fig. 1). As discussed in [9], the GNR is equivalent to a growing network with attachment rate $A_k = k+\lambda$, with $\lambda = r^{-1}-2$. Thus, for example, the GNR with $r=1/2$ corresponds to the growing network with linear preferential attachment, $A_k = k$. Simulation of the GNR is extremely simple because the selection of the initial target node is purely random and the ensuing re-direction step is local.

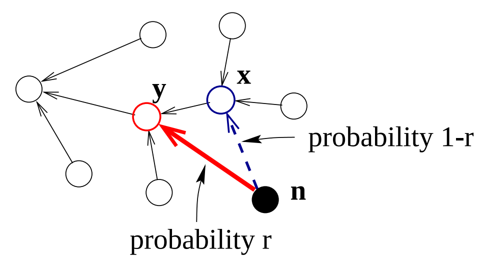

FIG. 1. The re-direction process. The new node $\mathbf{n}$ selects a random target node $\mathbf{x}$. With probability $1-r$ a link is established to this target node (dashed), while with probability r the link is established to $\mathbf{y}$, the ancestor of $\mathbf{x}$ (solid).

There is, however, an important subtlety about this equivalence that was not discussed previously in Ref. [9]. Namely, the redirection process does not apply when a node has no ancestor. By construction, every node that is added to the network does have a single ancestor, but some primordial nodes may have none. For example, for the very natural "dimer" initial condition ∘—∘, the seed node on the left has no ancestor and the GNR construction for this node is ambiguous. One way to resolve this dilemma is to adopt the "triangle" initial condition in which there are 3 nodes in a triangle with cyclic con-

nections between nodes. This leads to the correct attachment rate for each node for any value of λ. We therefore typically use this initial state to generate degree distribution data. On the other hand, theoretical analysis is simpler for the dimer initial condition. This state can also be simulated in a simple manner (for the case $\lambda = 0$) by a slightly modified GNR construction in which direct attachment to the seed node is not allowed. It is straightforward to check that this additional rule leads to the correct attachment rates for all the nodes in the network.

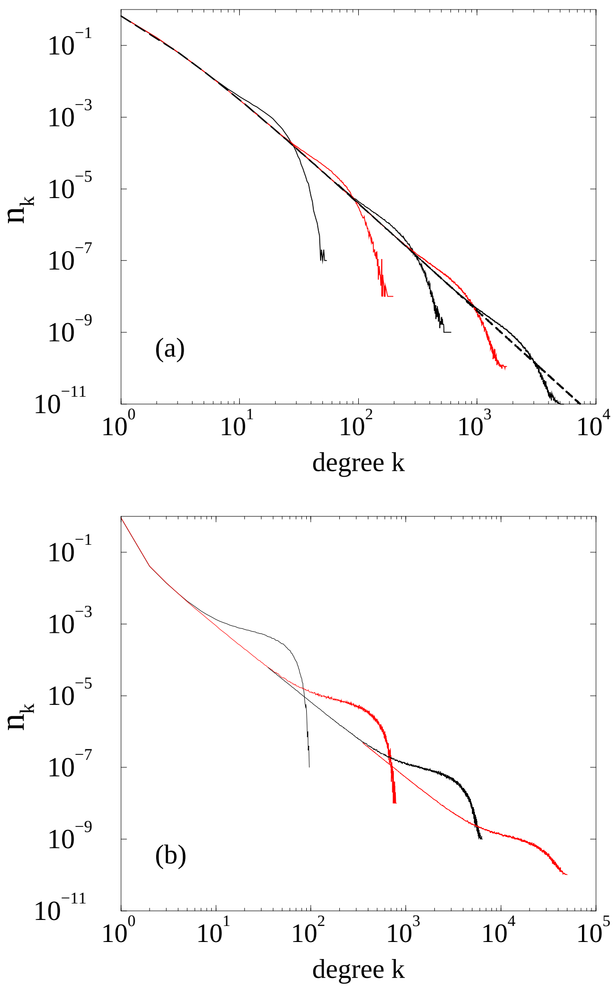

FIG. 2. Normalized degree distributions for the triangle initial condition for networks of $10^2, 10^3, \ldots$ links (upper left to lower right), with 10^5 realizations for each N, for (a) $A_k = k$ (up to 10^6 links) and (b) $A_k = k + \lambda$, with $\lambda = -0.9$ (up to 10^5 links). In (a), the dashed line is the asymptotic result $n_k = 4/[k(k+1)(k+2)]$; the last three data sets were averaged over 3, 9, and 27 points, respectively. In (b), the last two data sets were averaged over 10 and 100 points, respectively.

Figure 2 shows the average degree distribution for attachment rates $A_k = k$ and $A_k = k+\lambda$ with $\lambda = -0.9$ for the triangle initial condition. This latter value of λ gives results that are representative for values of λ close to -1. The data exhibits a shoulder at $k \approx k_{\max}$ that is much more pronounced when $\lambda < 0$ (Fig. 2(b)). This shoulder is also at odds with the natural expectation that the average degree distribution should exhibit a monotonic cutoff when k becomes of the order of $k_{\max}$. This shoulder turns into a clearly-resolved peak that exhibits relatively good data collapse when the degree distribution is re-expressed in the scaling form of Eq. (3) (Fig. 3). Conversely, the magnitude of the peak diminishes rapidly when λ is positive and becomes imperceptible for $\lambda \gtrsim 0.5$.

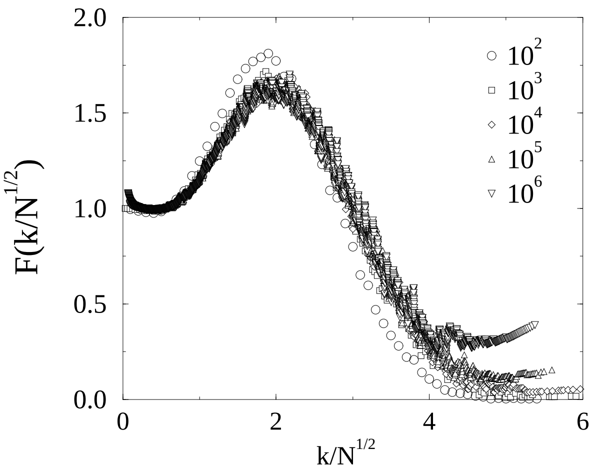

FIG. 3. The corresponding scaling function $F(\xi)$ in Eq. (3) for the data in Fig. 2(a).

In the following two sections, we will attempt to understand this anomalous feature of the degree distribution by studying the rate equations for the node degrees of finite networks.

IV. CONTINUUM FORMULATION

We focus on the case of the linear attachment rate $A_k = k$ and briefly quote corresponding results for other attachment rates. In the continuum approach, N is treated as continuously varying. Then the change in the average degree distribution satisfies the rate equation

$$\frac{d\,\langle N_k(N)\rangle}{dN} = \left\langle \frac{(k-1)N_{k-1}(N) - kN_k(N)}{2N} \right\rangle + \delta_{k,1}. \quad (4)$$

We assume the dimer initial state – two nodes connected by a single link so that $N_k(N=1) = 2\delta_{k,1}$.

Equations (4) are recursive and can be solved sequentially, starting with $\langle N_1\rangle$. Explicit results for $\langle N_k\rangle$, $k \le 4$, are given in Appendix A. These expressions show that the dominant contribution in the $N \to \infty$ limit is linear in N and this corresponds to the solution in Eq. (2). Indeed, if we substitute $\langle N_k(N)\rangle = n_k N$ into Eqs. (4), we obtain the recursion $n_k = n_{k-1}(k-1)/(k+2)$, whose solution is Eq. (2). From the first few $\langle N_k\rangle$, it is easy to see that the first correction to this leading behavior is of the order

of $N^{-1/2}$. Substituting $\langle N_k(N)\rangle = n_k N + A_k N^{-1/2}$ into Eqs. (4) and keeping the first two terms in each $\langle N_k\rangle$, we find $A_k = 4/3$. Continuing this procedure systematically, we arrive at the expansion:

$$
\begin{aligned}
\langle N_k(N)\rangle \; &= n_k N + \frac{4}{3}\,\frac{1}{N^{1/2}} - \frac{3}{2}\,\frac{k-1}{N} \\
&+ \frac{4}{5}\,\frac{(k-1)(k-2)}{N^{3/2}} \\
&- \frac{5}{18}\,\frac{(k-1)(k-2)(k-3)}{N^2} \\
&+ \frac{1}{14}\,\frac{(k-1)(k-2)(k-3)(k-4)}{N^{5/2}} + \ldots \qquad (5)
\end{aligned}
$$

In general, the right-hand side contains $k+1$ terms which can be written more succinctly as

$$
\langle N_k(N)\rangle = n_k N + \frac{1}{N^{1/2}} \sum_{j=0}^{k-1} \frac{\Gamma(k)}{\Gamma(k-j)}\,\frac{(-1)^j \nu_j}{N^{j/2}}. \qquad (6)
$$

The coefficients $\nu_j = (2j+4)/[j!(j+3)]$ may be obtained by imposing the initial condition $N_k(1) = 2\delta_{k,1}$ as each $\langle N_k\rangle$ is computed; a simpler way of obtaining these coefficients will be explained below. Note that expansion (5) is asymptotic because successive terms decrease only when $k \ll \sqrt{N}$.

A more convenient way to solve Eqs. (4) is in terms of the generating function

$$
\mathcal{N}(N,z) = \sum_{k=1}^{\infty} \langle N_k(N)\rangle\; z^k. \qquad (7)
$$

Multiplying Eq. (4) by z^k and summing over k, the generating function satisfies the following partial differential equation

$$
\left(2N\,\frac{\partial}{\partial N} + z(1-z)\,\frac{\partial}{\partial z}\right)\mathcal{N}(N,z) = 2Nz. \qquad (8)
$$

The initial condition is $\mathcal{N}(1,z) = 2z$, corresponding to a starting point of two nodes and a single connecting link.

We reduce Eq. (8) to a wave equation with constant coefficients by changing from the variables (N,z) to $(\ln\sqrt{N}, \ln[z/(1-z)])$. Then by introducing the rotated coordinates x, y such that $x + y = \ln\sqrt{N}$ and $x - y = \ln[z/(1-z)]$, we recast the wave equation into

$$
\frac{\partial \mathcal{N}(x,y)}{\partial x} = \frac{2\,e^{3x+2y}}{e^x + e^y}\,, \qquad (9)
$$

whose general solution is

$$
\mathcal{N}(x,y) = e^{2x+2y} - 2e^{x+3y} + 2e^{4y}\ln\left(e^x + e^y\right) + G(y).
$$

Finally, $G(y)$ is found by imposing the initial condition $\mathcal{N}(1,z) = 2z$. When $N = 1$, we have $x = -y$, so that the initial condition becomes $\mathcal{N}(-y,y) = 2/(1+e^{2y})$. Therefore

$$
G(y) = \frac{2}{1+e^{2y}} - 1 + 2e^{2y} - 2e^{4y}\ln\left(e^{-y} + e^y\right)
$$

and finally

$$
\begin{aligned}
\mathcal{N}(x,y) = e^{2y}\left(e^{2x} - 2e^{x+y} + 2\right) + \frac{1-e^{2y}}{1+e^{2y}} \\
+ 2e^{4y}\ln\left(\frac{e^{x+y} + e^{2y}}{1+e^{2y}}\right). \qquad (10)
\end{aligned}
$$

Using $e^{2y} = (z^{-1}-1)\sqrt{N}$ and $e^{x+y} = \sqrt{N}$, we re-express the generating function in term of the original variables

$$
\begin{aligned}
\mathcal{N}(N,z) = (3-2z^{-1})N + 2(z^{-1}-1)\sqrt{N} \\
+ \frac{1-(z^{-1}-1)\sqrt{N}}{1+(z^{-1}-1)\sqrt{N}} \\
- 2(z^{-1}-1)^2 N\,\ln\left(1 - z + \frac{z}{\sqrt{N}}\right). \qquad (11)
\end{aligned}
$$

We are primarily interested in the degree distribution for nodes whose degree is of the order of $k_{\max} \approx \sqrt{N}$. This part of the distribution can be extracted from the limiting behavior of the generating function $\mathcal{N}(N,z)$ as $z \to 1$ from below. Since the interesting range is $k \approx \sqrt{N}$, it is convenient to write

$$
z^{-1} = 1 + \frac{s}{\sqrt{N}} \qquad (12)
$$

and keep s finite while taking $N \to \infty$ limit. We simplify still further by eliminating the contribution to the generating function from the power-law tail of n_k in Eq. (2). For this purpose we consider the modified generating function

$$
\left(z^2\,\frac{\partial}{\partial z}\right)^3 \mathcal{N} = \sum_{k=1}^{\infty} (k+2)(k+1)k\,\langle N_k\rangle\; z^{k+3} \qquad (13)
$$

which is constructed so that the derivatives multiply the degree distribution by just the right factors to eliminate the power law tail. The leading behavior of this modified generating function will therefore provide the scaling function $F(\xi)$ of Eq. (3).

We now substitute Eq. (12) and the anticipated scaling form of Eq. (3) into the right-hand side of Eq. (13) and replace the sum by an integral. This gives the Laplace transform of the scaling function times a prefactor,

$$
4N^{3/2}\int_0^{\infty} d\xi\, F(\xi)\, e^{-\xi s}\,, \qquad (14)
$$

with $\xi = k/N^{1/2}$. Using Eq. (11), we compute the derivative on the left-hand side of Eq. (13). In the $N \to \infty$ limit, this derivative becomes $4N^{3/2}J(s)$ with

$$
J(s) = \frac{1}{1+s} + \frac{1}{(1+s)^2} + \frac{1}{(1+s)^3} + \frac{3}{(1+s)^4}\,. \qquad (15)
$$

This is just the Laplace transform of the scaling function. Inverting the Laplace transform then yields

$$F(\xi) = (1+\xi)\left(1+\frac{\xi^2}{2}\right)\, e^{-\xi}\,. \tag{16}$$

Notice that the coefficients ν_j in Eq. (6) can be obtained by expanding F in a Taylor series. This is a much simpler approach than solving each $\langle N_k(N)\rangle$ directly.

An important feature of the degree distribution is that it depends significantly on the initial condition. For example, for the triangle initial condition, solving Eq. (8) subject to $N_k^\Delta(N=3) = 3\delta_{k,2}$, or $\mathcal{N}^\Delta(3,z) = 3z^2$, yields

$$\begin{aligned}\mathcal{N}^\Delta(N,z) &= (3-2z^{-1})N + 2(z^{-1}-1)\sqrt{3N}\\ &\quad + 3\left(1+(z^{-1}-1)\sqrt{N/3}\right)^{-2} - 3\\ &\quad - 2(z^{-1}-1)^2 N\,\ln\left(1-z+z\sqrt{\frac{3}{N}}\right).\end{aligned} \tag{17}$$

Repeating the steps used to deduce the scaling function (16) from Eq. (11), we now find

$$F^\Delta(\xi) = \left(1+\eta+\frac{\eta^2}{2}+\frac{\eta^4}{4}\right)\, e^{-\eta}, \qquad \eta\equiv\xi\sqrt{3}. \tag{18}$$

Therefore small differences in the initial condition translate to discrepancies of the order of $\sqrt{N}$ in the degree distribution of a finite network of N links. Thus the properties of the nodes with the largest degrees are quite sensitive to the first few growth steps of the network (see also Ref. [6]).

While this initial condition dependence is real, there is also a spurious aspect to this effect. This may be illustrated by considering the linear trimer initial condition ◦—◦—◦. This is the unique outcome of the dimer initial condition after one node has been added. These two initial conditions should therefore lead to the same degree distribution. However, for the linear trimer initial state ($N_k(N=2) = 2\delta_{k,1}+\delta_{k,2}$) the continuum approach gives the scaling function,

$$F(\xi) = \left(1+\eta+\frac{\eta^2}{2}+\frac{\eta^3}{4}+\frac{\eta^4}{8}\right)\, e^{-\eta}, \qquad \eta\equiv\xi\sqrt{2},$$

which is distinct from Eq. (16)! This anomaly highlights one basic limitation of the continuum formulation.

Finally, we mention that parallel results can be obtained for the general case of the shifted linear attachment rate, $A_k = k+\lambda$. The rate equation for the average degree distribution is

$$\frac{d\,\langle N_k(N)\rangle}{dN} = \left\langle \frac{A_{k-1}N_{k-1}(N) - A_k N_k(N)}{A}\right\rangle + \delta_{k,1}\,,$$

where $A = \sum A_k N_k = \sum (k+\lambda)N_k$. To compute A we use the sum rules $\sum kN_k = 2N$ (every link contributes twice to the total degree), as well as $\sum N_k = N+1$ (for any tree initial condition) or $\sum N_k = N$ (for an initial condition that has the topology of a single cycle). To simplify final formulae, we use the latter topology (specifically, the triangle initial condition) so that $A = (2+\lambda)N$.

Solving the above rate equations successively, we find that the first two terms in the asymptotic series for $\langle N_k^\Delta(N)\rangle$ are

$$\left\langle N_k^\Delta(N)\right\rangle \sim n_k\, N + n_k'\, N^{-(1+\lambda)/(2+\lambda)} \tag{19}$$

with

$$\begin{aligned} n_k &= (2+\lambda)\,\frac{\Gamma(3+2\lambda)}{\Gamma(1+\lambda)}\,\frac{\Gamma(k+\lambda)}{\Gamma(k+3+2\lambda)}\,,\\ n_k' &= -\frac{2+\lambda}{3+2\lambda}\,\frac{3^{(3+2\lambda)/(2+\lambda)}}{\Gamma(1+\lambda)}\,\frac{\Gamma(k+\lambda)}{\Gamma(k)}\,.\end{aligned}$$

The corresponding leading behaviors are $n_k \propto k^{-(3+\lambda)}$ and $n_k' \propto k^\lambda$. Thus the two contributions to the degree distribution in Eq. (19) are comparable when $k \approx N^{1/(2+\lambda)}$. This value coincides with maximal degree $k_{\max}$ that is obtained by the extreme value condition $\sum_{k\geq k_{\max}} N/k^{3+\lambda} \approx 1$. Once again the degree distribution is described by a scaling function in the dimensionless variable $\xi = k/N^{1/(2+\lambda)}$.

V. DISCRETE APPROACH

We now turn now to the discrete approach for the network evolution. That is, one link is introduced at each discrete time step; this corresponds exactly to what occurs in the simulation. We again focus on the caes of the linear attachment rate $A_k = k$. We first treat in detail the case of nodes of degree one and then extend our approach to nodes of higher degrees. Finally, we give a scaling description for the degree distribution itself.

A. Nodes of Degree One

The number of nodes of degree one, $N_1(N)$, is a random variable that changes according to

$$N_1(N+1) = \begin{cases} N_1(N) & \text{prob. } \dfrac{N_1}{2N}\\ N_1(N)+1 & \text{prob. } 1-\dfrac{N_1}{2N}\end{cases} \tag{20}$$

after each node addition event. That is, with probability $N_1/2N$, a newly-introduced node attaches to a node of degree one; in this case, the number of nodes of degree one does not change. Conversely, with probability $(1-N_1/2N)$, the new node attaches to a node of degree greater than one and N_1 thus increases by one. Therefore

$$\begin{aligned}\langle N_1(N+1)\rangle &= \left\langle \frac{N_1^2(N)}{2N}\right\rangle\\ &\quad + \left\langle N_1(N) + 1 - \frac{N_1^2(N)}{2N} - \frac{N_1(N)}{2N}\right\rangle,\end{aligned}$$

from which

$$\langle N_1(N+1)\rangle = 1 + \left(1 - \frac{1}{2N}\right)\langle N_1(N)\rangle. \tag{21}$$

We take the initial condition $\langle N_1(1)\rangle = N_1(1) = 2$.

We solve this recursion in terms of the generating function $\mathcal{X}_1(w) = \sum_{N\geq 1}\langle N_1(N)\rangle\, w^{N-1}$. We therefore multiply Eq. (21) by Nw^{N-1} and sum over $N \geq 1$ to convert this recursion into the differential equation

$$\frac{d\mathcal{X}_1}{dw} = \frac{1}{(1-w)^2} + \frac{1}{2}\,\mathcal{X}_1 + w\,\frac{d\mathcal{X}_1}{dw}\,. \tag{22}$$

Solving Eq. (22) subject to the initial condition $\mathcal{X}_1(0) = 2$ gives

$$\mathcal{X}_1(w) = \frac{2}{3}\,\frac{1}{(1-w)^2} + \frac{4}{3}\,\frac{1}{(1-w)^{1/2}}\,. \tag{23}$$

Finally, we expand $\mathcal{X}_1(w)$ in a Taylor series in w to obtain

$$\langle N_1(N)\rangle = \frac{2}{3}\,N + \frac{4}{3\sqrt{\pi}}\,\frac{\Gamma\left(N-\frac{1}{2}\right)}{\Gamma(N)}\,. \tag{24}$$

The leading term is identical to that in the continuum approach (cf. Appendix A), but the coefficient of the correction term is $4/(3\sqrt{\pi}) \approx 0.7523$, compared to $4/3$ in the continuum approach.

The discrete approach is also suited to analyzing higher moments of the random variable $N_1(N)$. The second moment $\langle N_1^2(N)\rangle$ plays an especially important role as we can then obtain the variance $\sigma_1^2 = \langle N_1^2(N)\rangle - \langle N_1(N)\rangle^2$ and thereby quantify fluctuations. From Eq. (20) this second moment $\langle N_1^2(N)\rangle$ obeys the following recursion formula

$$\begin{aligned}\left\langle N_1^2(N+1)\right\rangle &= 1 + \left(1 - \frac{1}{N}\right)\langle N_1^2(N)\rangle \\ &\quad + \left(2 - \frac{1}{2N}\right)\langle N_1(N)\rangle.\end{aligned} \tag{25}$$

The solution to this recursion is outlined in Appendix B and the final result is

$$\begin{aligned}\langle N_1^2(N)\rangle &= \frac{4}{9}\,N(N+1) - \frac{1}{3}\,N + \frac{16}{9\sqrt{\pi}}\,\frac{\Gamma\left(N+\frac{1}{2}\right)}{\Gamma(N)} \\ &\quad - \frac{4}{3\sqrt{\pi}}\,\frac{\Gamma\left(N-\frac{1}{2}\right)}{\Gamma(N)} + \frac{35}{9}\,\delta_{N,1}.\end{aligned} \tag{26}$$

In the large N limit, we use Stirling's formula to give, for the variance,

$$\sigma_1^2 = \frac{N}{9} - \frac{20}{9\sqrt{\pi}}\,\frac{1}{N^{1/2}} - \frac{16}{9\sqrt{\pi}}\,\frac{1}{N} + \ldots \tag{27}$$

To obtain the entire probability distribution $P(N_1, N)$ one must solve

$$\begin{aligned}P(N_1, N+1) &= \frac{N_1}{2N}\,P(N_1, N) \\ &\quad + \left(1 - \frac{N_1 - 1}{2N}\right)P(N_1 - 1, N).\end{aligned} \tag{28}$$

By the Markov nature of the process, $P(N_1, N)$ should approach a Gaussian distribution in the large N limit. Numerically, we indeed find a Gaussian distribution with a peak at $2N/3$ and dispersion $\frac{1}{3}\sqrt{N}$ in agreement with our theoretical results for $\langle N_1(N)\rangle$ and $\langle N_1^2(N)\rangle$.

B. Degree Greater Than One

For $k \geq 2$, the random variable $N_k \equiv N_k(N)$ changes according to

$$N_k(N+1) = \begin{cases} N_k - 1 & \text{prob. } \dfrac{kN_k}{2N} \\[2ex] N_k + 1 & \text{prob. } \dfrac{(k-1)N_{k-1}}{2N} \\[2ex] N_k & \text{prob. } 1 - \dfrac{(k-1)N_{k-1} + kN_k}{2N} \end{cases} \tag{29}$$

at each node addition event. Again, because of the Markov nature of this process, we anticipate that $P(N_k, N)$ approaches a Gaussian distribution for every *fixed* degree k; therefore, we only need calculate $\langle N_k(N)\rangle$ and $\langle N_k^2(N)\rangle$ to infer the asymptotic distribution. To determine the first moment, we repeat the steps described in detail for $k = 1$ and obtain the recursion formula

$$\begin{aligned}\langle N_k(N+1)\rangle &= \langle N_k(N)\rangle \\ &\quad + \left\langle \frac{(k-1)N_{k-1}(N) - kN_k(N)}{2N}\right\rangle.\end{aligned} \tag{30}$$

The solution to this recursion is given in Appendix C and explicit formulae for $\langle N_k(N)\rangle$ for $k \leq 5$ are also quoted. Qualitatively, these results closely correspond to the asymptotic series for $\langle N_k(N)\rangle$ in the continuum formulation (Eq. (5)) but with somewhat different coefficients in the correction terms.

The determination of the second moment $\langle N_k^2\rangle$ is more complicated because it is coupled to $\langle N_{k-1}N_k\rangle$, which in turn is coupled to $\langle N_{k-2}N_k\rangle$, *etc.* However, we can still determine $\langle N_k^2\rangle$ for small k (Appendix D). From the structure of the rate equations, our general conclusion is that $\sigma_k^2 = \langle N_k^2(N)\rangle - \langle N_k(N)\rangle^2 = \mu_k N$. Therefore the distribution of $N_k(N)$ approaches a Gaussian for each k as $N \to \infty$.

C. Generating Function Approach

In close analogy with Sec. III, we now obtain the generating function for $\langle N_k(N)\rangle$, from which the exact scaling function in Eq. (3) can be deduced. Since Eq. (30) involves two discrete variables, k and N, it proves useful to introduce the two-variable generating function

$$\mathcal{N}(w,z)=\sum_{N=1}^{\infty}\sum_{k=1}^{\infty}\langle N_k(N)\rangle\, w^{N-1}\, z^k\,. \tag{31}$$

The governing equation for $\mathcal{N}(w,z)$ that is obtained from Eq. (30), is

$$\left(2(1-w)\,\frac{\partial}{\partial w}+z(1-z)\,\frac{\partial}{\partial z}-2\right)\mathcal{N}=\frac{2z}{(1-w)^2}\,. \tag{32}$$

This is similar to Eq. (8) and can be solved accordingly. We introduce the rotated variables x, y such that

$$x+y=-\frac{1}{2}\,\ln(1-w),\qquad x-y=\ln\frac{z}{1-z}, \tag{33}$$

to recast Eq. (32) into

$$\left(\frac{\partial}{\partial x}-2\right)\mathcal{N}(x,y)=\frac{2\,e^{5x+4y}}{e^x+e^y}. \tag{34}$$

The general solution is

$$\begin{aligned}\mathcal{N}(x,y)&=e^{4x+4y}-2e^{3x+5y}\\&\quad+2e^{2x+6y}\ln\left(e^x+e^y\right)+e^{2x}G(y),\end{aligned}$$

and the function $G(y)$ is found from the initial condition $\mathcal{N}(w=0,z)=2z$. When $w=0$, we have $x=-y=\frac{1}{2}\,\ln[z/(1-z0]$, and hence $\mathcal{N}(-y,y)=2/(1+e^{2y})$. Therefore

$$G(y)=\frac{2e^{2y}}{1+e^{2y}}-e^{2y}+2e^{4y}-2e^{6y}\ln\left(e^{-y}+e^{y}\right)\,,$$

and finally

$$\begin{aligned}\mathcal{N}(x,y)&=e^{4x+4y}-2e^{3x+5y}-e^{2x+2y}+2e^{2x+4y}\\&\quad+2\,\frac{e^{2x+2y}}{1+e^{2y}}+2e^{2x+6y}\ln\left(\frac{e^{x+y}+e^{2y}}{1+e^{2y}}\right).\end{aligned} \tag{35}$$

In term of the original w, z variables,

$$\begin{aligned}\mathcal{N}(w,z)&=\frac{(3-2z^{-1})}{(1-w)^2}-\frac{1}{1-w}\\&\quad+\frac{2(z^{-1}-1)}{(1-w)^{3/2}}+\frac{2(1-w)^{-1/2}}{(z^{-1}-1)+(1-w)^{1/2}}\\&\quad-\frac{2(z^{-1}-1)^2}{(1-w)^2}\,\ln\left[1-z+z(1-w)^{1/2}\right].\end{aligned} \tag{36}$$

By expanding $\mathcal{N}(w,z)$, we can in principle determine all the $\langle N_k(N)\rangle$.

D. Scaling Function

To extract the scaling function $F(\xi)$ from the generating function $\mathcal{N}(w,z)$ we use the same approach as in Sec. IV. The details are given in Appendix E and the final result is

$$F(\xi)=\text{erfc}\left(\frac{\xi}{2}\right)+\frac{2\xi+\xi^3}{\sqrt{4\pi}}\;e^{-\xi^2/4}\,, \tag{37}$$

where $\text{erfc}(x)$ is the complementary error function. A similar result for a related network model was found previously by Dorogovtsev et al. [14]. Notice that the exact form for $F(\xi)$ vanishes much more quickly than predicted by the continuum approach. When $k\gg\sqrt{N}$, the continuum approach gives

$$\langle N_k(N)\rangle_{\text{cont.}}\to\frac{2}{\sqrt{N}}\;e^{-k/\sqrt{N}}\,, \tag{38}$$

while the exact average degree distribution has a Gaussian large-degree tail

$$\langle N_k(N)\rangle_{\text{exact}}\to\frac{2}{\sqrt{\pi N}}\;e^{-k^2/4N}\,, \tag{39}$$

The scaling function in Eq. (37) quantitatively accounts for the shoulder in the degree distribution. In contrast, while the scaling function from the continuum approach does exhibit a peak, it is both quantitatively and qualitatively inaccurate (Fig. 4).

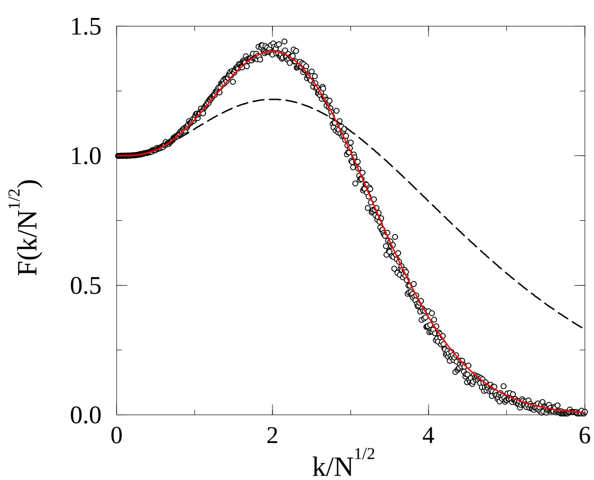

FIG. 4. Comparison between the scaling function $F(\xi)$, with $\xi=k/N^{1/2}$, in the continuum approximation [Eq. (18), dashed curve] and in the discrete approach [Eq. (37), solid curve]. The circles give the simulation data of 10^6 realizations of a network with $N=10^4$ links for the dimer initial condition; these data coincide with the theoretical prediction.

VI. HIGHER MOMENTS AND THEIR FLUCTUATION

We now turn to higher moments of the degree distribution, as well as the fluctuation in these quantities between different realizations of the network. While the zeroth and first moments of the degree distribution are simply related to the total number of links for *any* network topology, the higher moments are not so simply characterized, but instead reflect the power-law tail of the degree distribution.

We first compare the moments of the average degree distribution to appreciate the difference between the continuum and exact descriptions. For the second moment, we use the identity

$$\sum_{k=1}^{\infty} k(k+1)\langle N_k\rangle \equiv \left(z^2\frac{\partial}{\partial z}\right)^2 \mathcal{N}(N,z)\Big|_{z=1}. \tag{40}$$

Using $\mathcal{N}(N,z)$ from Eq. (11), together with the value of the first moment, we obtain, in the continuum approximation,

$$\langle k^2\rangle_{\rm cont.} \equiv \sum_{k=1}^{\infty} k^2\langle N_k\rangle_{\rm cont.} = 2N\ln N + 2N. \tag{41}$$

On the other hand, using the exact discrete expression (36) we find

$$\left(z^2\frac{\partial}{\partial z}\right)^2 \mathcal{N}(w,z)\Big|_{z=1} = \frac{4-2\ln(1-w)}{(1-w)^2}\,,$$

which we then expand in a series in w to yield, for the second moment,

$$\langle k^2\rangle_{\rm exact} \equiv \sum_{k=1}^{\infty} k^2\langle N_k\rangle_{\rm exact} = 2NH_N\,. \tag{42}$$

Here $H_N = \sum_{1\le j\le N} j^{-1}$ is the harmonic number [16]. In the large N limit, therefore,

$$\langle k^2\rangle_{\rm exact} = 2N\ln N + 2\gamma N + 1 - \frac{1}{6N} + \ldots\,,$$

where $\gamma \cong 0.5772166$ is Euler's constant.

For higher moments, even the leading term given by the continuum approach is erroneous. For example,

$$\langle k^3\rangle_{\rm cont.} = 24N^{3/2} - 6N\ln N - 22N\,, \tag{43}$$

while the exact value is

$$\langle k^3\rangle_{\rm exact} = \frac{32}{\sqrt{\pi}}\,\frac{\Gamma\left(N+\frac{3}{2}\right)}{\Gamma(N)} - 6NH_N - 16N\,. \tag{44}$$

More generally, the dependence of the moments on N stems from the power-law tail of the degree distribution $\langle N_k\rangle \propto N/k^3$. From this asymptotic distribution, a suitably normalized set of measures for the mean degree

$$\mathcal{M}_n = \left(\frac{\langle k^n\rangle}{\langle k^0\rangle}\right)^{1/n}, \tag{45}$$

has the following N dependence:

$$\mathcal{M}_n \propto \begin{cases} \text{const.} & n<2 \\ \ln N & n=2 \\ N^{(n-2)/2} & n>2 \end{cases} \tag{46}$$

In a related vein, we also study the fluctuations in these moments between different realizations of the network growth. That is, we record the value of $\langle k^2\rangle$ for each realization of the network to obtain the underlying distribution $P(\langle k^2\rangle)$. A typical result is shown in Fig. 5. Notice that the distribution of $\langle k^2\rangle$ is relatively broad with an exponential tail. The distributions of higher moments are even broader, with each being dominated by the realizations with the largest value of the corresponding moment.

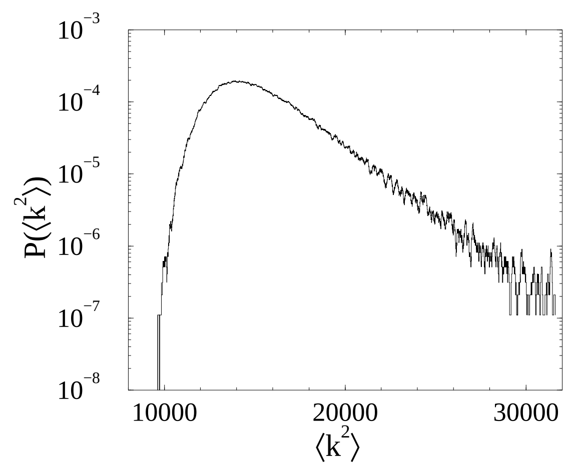

FIG. 5. Distribution of $\langle k^2\rangle$ for 10^5 realizations of a growing network with $N = 10^3$ for attachment rate $A_k = k$ with the triangle initial condition. The raw data has been smoothed over a 100-point range.

VII. CONCLUDING REMARKS

We studied the role of finiteness on the degree distributions of growing networks with a node attachment rate of the form $A_k = k+\lambda$. For finite networks, fluctuations are no longer negligible and a stochastic approach is needed to analyze these properties. We found the average degree distribution within an approximate continuum formulation and by an exact discrete approach. The continuum approach has the advantage of being much simpler than the discrete formulation, but does not provide a quantitatively accurate description of the large-k tail of the degree distribution.

We also argued that the degree distribution $N_k(N)$, when considered as the random variable in k, exhibits self

averaging, *i.e.*, the relative fluctuations in $N_k(N)$ diminish as $N \to \infty$. Moreover, the variance $\sigma_k^2 = \langle N_k^2\rangle - \langle N_k\rangle^2$ varies linearly with N, and the probability distribution $P(N_k, N)$ approaches a Gaussian. To support these assertions, we computed σ_k^2 for $k = 1, 2$. These partial results support our general hypothesis that fluctuations in $N_k(N)$ are Gaussian. Perhaps the Van Kampen Ω-expansion [17] would prove to be a more appropriate analysis tool to undertake a systematic study of fluctuations in growing networks.

Of course, the random variables $N_k(N)$ should be Gaussian only for sufficiently small k, *viz.*, as long as $\langle N_k\rangle \gg 1$, or equivalently, $k \ll N^{1/(3+\lambda)}$. On the other hand, fluctuations become large and non-Gaussian when $k \propto N^{1/(3+\lambda)}$. Determining the fluctuations in this degree range seems to be difficult, as one must study the master equation for the joint probability distribution.

In this work, we limited ourselves to the degree distribution; this is perhaps the most important and also the most easily analyzable local structural characteristic of a network. However, recent investigations of growing networks has increasingly focused on global characteristics, such as the size distribution of connected components, see *e.g.*, Refs. [18–21]. The methods described in this paper should be applicable to probing fluctuations of the component size distribution and other global network characteristics. This direction seems especially exciting since the simplest growing network models that allow for a multiplicity of clusters exhibit a very unusual infinite-order percolation transition [18–21]. Thus one might anticipate interesting giant fluctuations near the percolation transition of these models.

We are grateful to NSF grant DMR9978902 for partial financial support of this research.

APPENDIX A: THE AVERAGE DEGREE DISTRIBUTION IN THE CONTINUUM FORMULATION

Within the continuum framework, the average degree distribution is described by Eqs. (4). Successively solving these equations by elementary methods, we obtain $\langle N_k(N)\rangle$. For $k = 1, 2, 3$, and 4 we obtain:

$$\begin{aligned}
\langle N_1(N)\rangle &= \frac{2}{3}\,N + \frac{4}{3}\,\frac{1}{N^{1/2}},\\
\langle N_2(N)\rangle &= \frac{1}{6}\,N + \frac{4}{3}\,\frac{1}{N^{1/2}} - \frac{3}{2}\,\frac{1}{N},\\
\langle N_3(N)\rangle &= \frac{1}{15}\,N + \frac{4}{3}\,\frac{1}{N^{1/2}} - 3\,\frac{1}{N} + \frac{8}{5}\,\frac{1}{N^{3/2}},\\
\langle N_4(N)\rangle &= \frac{1}{30}\,N + \frac{4}{3}\,\frac{1}{N^{1/2}} - \frac{9}{2}\,\frac{1}{N} + \frac{24}{5}\,\frac{1}{N^{3/2}} - \frac{5}{3}\,\frac{1}{N^2}.
\end{aligned}$$

APPENDIX B: GENERATING FUNCTION FOR $\langle N_1^2(N)\rangle$

To determine $\langle N_1^2(N)\rangle$, we introduce the generating function $\mathcal{Y}_1(w) = \sum_{N\ge1}\langle N_1^2(N)\rangle\, w^{N-1}$. This converts the recursion relation Eq. (25) into the differential equation for the generating function

$$(1-w)\,\frac{d\mathcal{Y}_1}{dw} = \frac{1}{(1-w)^2} + \frac{3}{2}\,\mathcal{X}_1 + 2w\,\frac{d\mathcal{X}_1}{dw}\,, \tag{B1}$$

with $\mathcal{X}_1(w)$ given by Eq. (23). Solving (B1) subject to the initial condition $\mathcal{Y}_1(0) = 4$ we obtain

$$\begin{aligned}
\mathcal{Y}_1(w) = &\frac{8}{9}\,\frac{1}{(1-w)^3} - \frac{1}{3}\,\frac{1}{(1-w)^2} + \frac{8}{9}\,\frac{1}{(1-w)^{3/2}}\\
&- \frac{4}{3}\,\frac{1}{(1-w)^{1/2}} + \frac{35}{9}\,,
\end{aligned} \tag{B2}$$

Expanding this generating function in a Taylor series then yields the result for $\langle N_1^2(N)\rangle$ quoted in Eq. (26).

APPENDIX C: GENERATING FUNCTION FOR FIRST MOMENT

Here we solve the recursion formula Eq. (30) for $\langle N_k(N)\rangle$. We first introduce the generating function $\mathcal{X}_k(w) = \sum_{N=1}^{\infty}\langle N_k(N)\rangle\, w^{N-1}$ to eliminate the variable N and convert Eq. (30) into a differential equation that relates $\mathcal{X}_k$ and $\mathcal{X}_{k-1}$. This equation is further simplified by making the transformation

$$\mathcal{X}_k(w) = (1-w)^{\frac{k}{2}-1}\,\mathcal{U}_k(u), \quad u = \frac{1}{\sqrt{1-w}} - 1. \tag{C1}$$

The resulting equation is

$$\frac{d\mathcal{U}_k}{du} = (k-1)\,\mathcal{U}_{k-1}, \qquad k \ge 2. \tag{C2}$$

Rewriting our previous solution (23) as

$$\mathcal{U}_1(u) = \frac{2}{3}\,u^3 + 2u^2 + 2u + 2, \tag{C3}$$

one can solve Eqs. (C2) subject to the initial condition $\mathcal{U}_k(u=0) = 0$ for $k \ge 2$. The final result is

$$\mathcal{U}_k(u) = \frac{4u^{k+2}}{k(k+1)(k+2)} + \frac{4u^{k+1}}{k(k+1)} + \frac{2u^k}{k} + 2u^{k-1}\,.$$

Using the binomial formula, we transform $\mathcal{X}_k(z)$ into the series

$$\begin{aligned}
\mathcal{X}_k(w) = &\frac{4}{k(k+1)(k+2)}\,\frac{1}{(1-w)^2} + \frac{4}{3}\,\frac{1}{(1-w)^{1/2}}\\
&+ 2\sum_{a=1}^{k-1}(-1)^a\,\frac{a+2}{a+3}\binom{k-1}{a}(1-w)^{(a-1)/2}\,.
\end{aligned}$$

Expanding $\mathcal{X}_k(w)$ in a Taylor series in w we obtain $\langle N_k(N)\rangle$. The analytic expressions for $\langle N_k(N)\rangle$ with

$k \le 5$ are obtained by expanding $\mathcal{X}_k(w)$ in a Taylor series. This gives

$$\begin{aligned}
\langle N_1(N)\rangle &= \frac{2}{3}\,N + \frac{4}{3\sqrt{\pi}}\;\frac{\Gamma\left(N-\frac{1}{2}\right)}{\Gamma(N)},\\
\langle N_2(N)\rangle &= \frac{1}{6}\,N + \frac{4}{3\sqrt{\pi}}\;\frac{\Gamma\left(N-\frac{1}{2}\right)}{\Gamma(N)} - \frac{3}{2}\,\delta_{N,1},\\
\langle N_3(N)\rangle &= \frac{1}{15}\,N + \frac{4}{3\sqrt{\pi}}\;\frac{\Gamma\left(N-\frac{1}{2}\right)}{\Gamma(N)} - 3\,\delta_{N,1}\\
&\quad - \frac{4}{5\sqrt{\pi}}\;\frac{\Gamma\left(N-\frac{3}{2}\right)}{\Gamma(N)},\\
\langle N_4(N)\rangle &= \frac{1}{30}\,N + \frac{4}{3\sqrt{\pi}}\;\frac{\Gamma\left(N-\frac{1}{2}\right)}{\Gamma(N)} - \frac{9}{2}\,\delta_{N,1}\\
&\quad - \frac{12}{5\sqrt{\pi}}\;\frac{\Gamma\left(N-\frac{3}{2}\right)}{\Gamma(N)} - \frac{5}{3}\,\delta_{N,1} + \frac{5}{3}\,\delta_{N,2}\\
\langle N_5(N)\rangle &= \frac{2}{105}\,N + \frac{4}{3\sqrt{\pi}}\;\frac{\Gamma\left(N-\frac{1}{2}\right)}{\Gamma(N)} - 6\,\delta_{N,1}\\
&\quad - \frac{24}{5\sqrt{\pi}}\;\frac{\Gamma\left(N-\frac{3}{2}\right)}{\Gamma(N)} - \frac{20}{3}\,\delta_{N,1} + \frac{20}{3}\,\delta_{N,2}\\
&\quad + \frac{9}{7\sqrt{\pi}}\;\frac{\Gamma\left(N-\frac{5}{2}\right)}{\Gamma(N)}\,.
\end{aligned}$$

Generally, there are slightly different formulae for even

$$\begin{aligned}
\langle N_{2k}(N)\rangle &= n_{2k}\,N + \sum_{j=1}^{k} A_{kj}\,\delta_{Nj}\\
&\quad + \sum_{i=0}^{k-1} \frac{4(i+1)}{2i+3}\binom{2k-1}{2i}\frac{\Gamma\left(N-\frac{1}{2}-i\right)}{\Gamma\left(\frac{1}{2}-i\right)\,\Gamma(N)}
\end{aligned}$$

and odd

$$\begin{aligned}
\langle N_{2k+1}(N)\rangle &= n_{2k+1}\,N + \sum_{j=1}^{k} B_{kj}\,\delta_{Nj}\\
&\quad + \sum_{i=0}^{k} \frac{4(i+1)}{2i+3}\binom{2k}{2i}\frac{\Gamma\left(N-\frac{1}{2}-i\right)}{\Gamma\left(\frac{1}{2}-i\right)\,\Gamma(N)}
\end{aligned}$$

indices. Here the n_k are given by Eqs. (2) and explicit expressions for the coefficients A_{kj} and B_{kj} could be found by expanding the polynomials in the generating functions $\mathcal{X}_{2k}(w)$ and $\mathcal{X}_{2k+1}(w)$.

APPENDIX D: HIGHER MOMENTS

Starting from Eq. (29), a straightforward computation yields

$$\begin{aligned}
\langle N_k^2\rangle &= \left(1-\frac{k}{N}\right)\langle N_k^2\rangle + \frac{k-1}{N}\,\langle N_{k-1}N_k\rangle\\
&\quad + \left\langle \frac{(k-1)N_{k-1}+kN_k}{2N}\right\rangle\,, \qquad \text{(D1)}
\end{aligned}$$

where the correlation function on the left-hand side is a function of $N+1$ and those on the right-hand side are functions of N. Obviously, $\langle N_k^2\rangle$ is coupled with $\langle N_{k-1}N_k\rangle$. The recursion relation for this correlation function reads (for $k \ge 3$)

$$\begin{aligned}
\langle N_{k-1}N_k\rangle &= \left(1-\frac{2k-1}{2N}\right)\langle N_{k-1}N_k\rangle + \frac{k-1}{2N}\,\langle N_{k-1}^2\rangle\\
&\quad + \frac{k-2}{2N}\,\langle N_{k-2}N_k\rangle - \frac{k-1}{2N}\langle N_{k-1}\rangle. \qquad \text{(D2)}
\end{aligned}$$

Fortunately no higher-order correlation functions appear, and additionally the total index decreases, *i.e.*, $\langle N_k^2\rangle$, whose total index is $2k$, involves the correlation function $\langle N_{k-1}N_k\rangle$, whose total index is $2k-1$. One therefore can determine all correlation functions by starting from the smallest total index and then working up to larger indices. For example, the first non-trivial correlation function $\langle N_1N_2\rangle$ whose total index equals three satisfies an equation slightly different from the general form of Eq. (D2), *viz.*,

$$\begin{aligned}
\langle N_1N_2\rangle &= \left(1-\frac{3}{2N}\right)\langle N_1N_2\rangle\\
&\quad + \frac{1}{2N}\,\langle N_1^2\rangle + \left(1-\frac{1}{N}\right)\langle N_2\rangle\,. \qquad \text{(D3)}
\end{aligned}$$

Notice here that we already know $\langle N_1^2\rangle$.

We can solve for $\langle N_1N_2\rangle$ using the generating function technique. We define the generating function $\mathcal{Z}_1(w) = \sum_{N\ge 1}\langle N_1(N)N_2(N)\rangle\, w^{N-1}$ which satisfies the differential equation

$$2(1-w)\,\frac{d\mathcal{Z}_1}{dw} = -\mathcal{Z}_1 + \mathcal{Y}_1 + 2w\,\frac{d\mathcal{X}_2}{dw}, \qquad \text{(D4)}$$

with solution

$$\begin{aligned}
\mathcal{Z}_1(w) &= \frac{2}{9}\,\frac{1}{(1-w)^3} - \frac{1}{5}\,\frac{1}{(1-w)^2} + \frac{5}{9}\,\frac{1}{(1-w)^{3/2}}\\
&\quad - \frac{4}{3}\,\frac{1}{(1-w)^{1/2}} - \frac{47}{15}\,(1-w)^{1/2} + \frac{35}{9}\,. \qquad \text{(D5)}
\end{aligned}$$

Expanding $\mathcal{Z}_1(w)$ in a power series in w we obtain

$$\begin{aligned}
\langle N_1N_2\rangle &= \frac{1}{9}\,N(N+1) - \frac{1}{5}\,N + \frac{10}{9\sqrt{\pi}}\;\frac{\Gamma\left(N+\frac{1}{2}\right)}{\Gamma(N)}\\
&\quad - \frac{4}{3\sqrt{\pi}}\;\frac{\Gamma\left(N-\frac{1}{2}\right)}{\Gamma(N)} + \frac{47}{30\sqrt{\pi}}\;\frac{\Gamma\left(N-\frac{3}{2}\right)}{\Gamma(N)}\\
&\quad + \frac{35}{9}\,\delta_{N,1}\,.
\end{aligned}$$

Asymptotically, $\langle N_1N_2\rangle \to \langle N_1\rangle\langle N_2\rangle$ as expected.

There are two correlation functions, $\langle N_2^2\rangle$ and $\langle N_1N_3\rangle$, whose total index equals four. The former satisfies Eq. (D1) with $k=2$, *i.e.*,

$$\langle N_2^2\rangle = \left(1-\frac{2}{N}\right)\langle N_2^2\rangle + \left\langle\frac{N_1+2N_2+2N_1N_2}{2N}\right\rangle ,$$

from which we determine the generating function

$$\mathcal{Y}_2(w) = \frac{1}{18}\,\frac{1}{(1-w)^3} + \frac{1}{10}\,\frac{1}{(1-w)^2} + \frac{2}{9}\,\frac{1}{(1-w)^{3/2}} + \frac{4}{9}\,\frac{1}{(1-w)^{1/2}} - \frac{94}{15}\,(1-w)^{1/2} + \frac{49}{9} - \frac{55}{18}\,w\,.$$

Expanding $\mathcal{Y}_2(w)$ we obtain

$$\langle N_2^2(N)\rangle = \frac{1}{36}\,N(N+1) + \frac{1}{10}\,N + \frac{4}{9\sqrt{\pi}}\,\frac{\Gamma\left(N+\frac{1}{2}\right)}{\Gamma(N)} + \frac{4}{9\sqrt{\pi}}\,\frac{\Gamma\left(N-\frac{1}{2}\right)}{\Gamma(N)} + \frac{47}{15\sqrt{\pi}}\,\frac{\Gamma\left(N-\frac{3}{2}\right)}{\Gamma(N)} + \frac{49}{9}\,\delta_{N,1} - \frac{55}{18}\,\delta_{N,2}\,.$$

In the large N limit, we find that variance grows linearly with N according to $\sigma_2^2 \sim \frac{23}{180}\,N$. It appears that

$$\sigma_k^2 \to \mu_k\,N \qquad \text{as} \quad N\to\infty, \tag{D6}$$

for all k, although we solved only the cases $k=1$ and 2, where $\mu_1 = \frac{1}{9}$ and $\mu_2 = \frac{23}{180}$.

APPENDIX E: SCALING FUNCTION IN THE DISCRETE APPROACH

To extract the scaling function from the generating function $\mathcal{N}(w,z)$ we adapt the technique employed in Sec. IV for discrete variables. We first write

$$z^{-1} = 1 + s\,\sqrt{1-w} \tag{E1}$$

and keep s finite while taking the $w\to 1$ limit. We again consider the modified generating function

$$\left(z^2\,\frac{\partial}{\partial z}\right)^3 \mathcal{N} = \sum_{N=1}^{\infty}\sum_{k=1}^{\infty} 4NF\left(\frac{k}{\sqrt{N}}\right)\,w^{N-1}z^{k+3}\,. \tag{E2}$$

On the right-hand side of this equation we have already replaced $(k+2)(k+1)k\langle N_k(N)\rangle$ by $4NF(k/\sqrt{N})$ as implied by Eqs. (2)–(3).

Substituting the exact expression (36) for the generating function into the left-hand side of Eq. (E2) and keeping only the dominant contribution gives

$$4(1-w)^{-5/2}J(s), \tag{E3}$$

with $J(s)$ given by Eq. (15). To simplify the right-hand side of Eq. (E2) we substitute Eq. (E1) and replace the sums by integrals. The dominant contribution in the $w\to 1$ limit is

$$4(1-w)^{-5/2}\int_0^\infty d\xi\, e^{-\xi s}\int_0^\infty d\eta\,\eta\,F(\xi\eta^{-1/2})\,e^{-\eta}\,, \tag{E4}$$

where $\xi = k\sqrt{1-w}$ and $\eta = N(1-w)$. Therefore the double integral in Eq. (E4) is equal to $J(s)$. The double integral can be interpreted as the Laplace transform $\hat\Phi(s) = \int_0^\infty d\xi\,\exp(-s\xi)\Phi(\xi)$ of the function

$$\Phi(\xi) = \int_0^\infty d\eta\,\eta\,F(\xi\eta^{-1/2})\,e^{-\eta}\,. \tag{E5}$$

We already know how to solve $\hat\Phi(s) = J(s)$, so

$$\Phi(\xi) = (1+\xi)\left(1+\frac{\xi^2}{2}\right)e^{-\xi}\,. \tag{E6}$$

To determine $F(\xi)$, we must solve the integral equation (E6) with $\Phi(\xi)$ given by Eq. (E5). To solve this integral equation, notice that $\Phi(\xi)$ is almost a Laplace transform of function F. Indeed, if instead of η and $F(\xi\eta^{-1/2})$ we use ζ and $G(\zeta)$ defined according to

$$\zeta = \frac{\eta}{\xi^2}, \qquad G(\zeta) = \zeta\,F(\zeta^{-1/2})\,, \tag{E7}$$

then we obtain $\Phi(\xi) = p^2\hat G(p)$, with $p=\xi^2$ being the Laplace variable and $\hat G(p) = \int_0^\infty d\zeta\,G(\zeta)\,\exp(-p\zeta)$. Rewriting the integral equation (E6) in terms of p gives

$$\hat G(p) = \left(p^{-2} + p^{-3/2} + \frac{1}{2}\,p^{-1} + \frac{1}{2}\,p^{-1/2}\right)\exp(-\sqrt{p})\,.$$

Inverting this Laplace transform yields [22]

$$G(\zeta) = \zeta\,\mathrm{erfc}\left(\frac{1}{\sqrt{4\zeta}}\right) + \frac{2\zeta+1}{\sqrt{4\pi\zeta}}\,e^{-1/4\zeta}\,, \tag{E8}$$

where $\mathrm{erfc}(x) = \frac{2}{\sqrt{\pi}}\int_x^\infty dt\,\exp(-t^2)$ is the complementary error function. Since $F(\xi) = \xi^2 G(\xi^{-2})$, see Eq. (E7), we arrive at the scaled average degree distribution quoted in Eq. (37).

[1] A. Broder *et al.*, Computer Networks, **33**, 309 (2000).
[2] W. Willinger *et al.*, Proc. Natl. Acad. Sci. **99**, 2573 (2002).
[3] A short review of applications of the rate equation approach to network growth is given in P. L. Krapivsky and S. Redner, Computer Networks, **39**, 262 (2002).
[4] It was also suggested that fluctuations can affect the growth of the Web itself, see e.g. B. A. Huberman and L. A. Adamic, Nature **401**, 131 (1999).
[5] A. A. Moreira, J. S. Andrade, and L. A. N. Amaral, *cond-mat*/0205411.
[6] P. L. Krapivsky and S. Redner, *cond-mat*/0207370.
[7] H. A. Simon, Biometrica **42**, 425 (1955); Infor. Control **3**, 80 (1960).

[8] A.-L. Barabási and R. Albert, Science **286**, 509 (1999).
[9] P. L. Krapivsky and S. Redner, Phys. Rev. E **63**, 066123 (2001).
[10] P. L. Krapivsky, S. Redner, and F. Leyvraz, Phys. Rev. Lett. **85**, 4629 (2000).
[11] S. N. Dorogovtsev, J. F. F. Mendes, and A. N. Samukhin, Phys. Rev. Lett. **85**, 4633 (2000).
[12] D. H. Zanette and S. C. Manrubia, Physica A **295**, 1 (2001).
[13] L. Kullmann and J. Kertész, Phys. Rev. E **63**, 051112 (2001).
[14] S. N. Dorogovtsev J. F. F. Mendes, and A. N. Samukhin, Phys. Rev. E **63**, 062101 (2001).
[15] Z. Burda, J. D. Correia, and A. Krzywicki, Phys. Rev. E **64**, 046118 (2001).
[16] R. L. Graham, D. E. Knuth, and O. Patashnik, *Concrete Mathematics: A Foundation for Computer Science* (Reading, Mass.: Addison-Wesley, 1989).
[17] N. G. Van Kampen, *Stochastic Processes in Physics and Chemistry* (North-Holland: Amsterdam, 1981).
[18] D. S. Callaway *et al.*, Phys. Rev. E **64**, 041902 (2001).
[19] S. N. Dorogovtsev, J. F. F. Mendes, and A. N. Samukhin, Phys. Rev. E **64**, 066110 (2001).
[20] J. Kim, P. L. Krapivsky, B. Kahng, and S. Redner, *cond-mat*/0203167.
[21] M. Bauer and D. Bernard, *cond-mat*/0203232.
[22] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions* (Dover Publications Inc., New York, 1970).